DOCUMENT RESUME

ED 141 388 TM 006 332

AUTHOR Haertel, Edward; Prescott, Suzanne
TITLE Observations of Teacher Behavior: A Problem of Two Random Effects.
PUB DATE [Apr 77]
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Analysis of Variance; Classroom Observation Techniques; *Data Analysis; Elementary School Teachers; *Research Design; Sampling; Secondary School Teachers; *Statistical Analysis; *Teacher Behavior
IDENTIFIERS *Random Effects

ABSTRACT

Several methods of analyzing random samples of teacher behavior are discussed. Their assumptions and sensitivity, and the inferences they permit, are compared. Data of this kind might be obtained whenever program implementation is assessed by monitoring specific teacher behaviors. The quasi F ratio can be used in analyzing such data, but error rates are difficult to estimate and the generalization of the quasi F ratio to multivariate observations is difficult. Two alternatives are presented: (1) Observations within teachers may be averaged, leaving teacher as the only random effect. (2) Teacher may be treated as a fixed effect, but this limits the kinds of inferences which may be drawn. Results of analyses by both methods are presented. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED141388

TM006 332

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Observations of Teacher Behavior: A Problem of Two Random Effects

Edward Haertel
University of Chicago

Suzanne Prescott
University of Chicago

Statisticians have long distinguished between fixed and random effects in analysis of variance (Wilk and Kempthorne, 1955). Unfortunately, educational researchers are often accustomed to assuming that all effects except Subject are fixed, an error which may lead to faulty statistical generalizations. In this paper two possible sources of additional random effects are described, and appropriate statistical procedures are presented.

The distinction between fixed and random effects is usually clearcut. If specific levels of a design variable are selected, the variable defines a fixed effect; if the levels to be included are randomly sampled from a large population of possible levels, the variable defines a random effect. Analysis of variance permits generalizations about the chosen levels of fixed effects across the randomly sampled dimensions. Ordinarily, replications are random and nested within all other effects, and all other effects are fixed. Inferences in this case are made concerning the effects of the specific treatment levels in the population from which the sample used in the experiment was drawn.

It is argued in this paper that in exploratory data analysis valuable information may be obtained by deliberately blurring this important distinction between fixed and random effects, treating random effects as fixed, provided that the limitations which this places upon the inference is kept in mind. At the same time we shall emphasize the importance of recognizing all random effects and treating them appropriately in confirmatory analyses.

For purposes of discussion, a small data set was extracted from a larger study, and analyzed according to some of the procedures to be described. Five elementary and junior high school teachers who agreed to cooperate were equipped with a telephone paging device which was sounded at random times throughout the school day. Each time the device beeped, the teacher was to complete a brief checklist describing the activity currently being performed and present affective state. Of the responses obtained, thirty-five were taken in the classroom. These were classified according to the activity in which the teacher was engaged: Group instruction, Individual instruction, and Maintainance activities, such as passing out papers, taking attendence, or supervising while a movie was being shown. Not all teachers were paged during all three activities from all three categories. It should be noted that this study was not a true experiment. No random assignment was made, and no causal inferences regarding activity and mood are justified by these data. The purpose of the study

was to detect differences in mood associated with different activities, not to show that engaging in different activities caused changes in mood.

The unequal cell sizes and empty cells in this set of data raise problems which are not central to the issue of fixed and random effects. For clarity, let us consider an ideal, balanced design along the same lines. Suppose that each of five teachers were paged and responded many times, and that, for each teacher, three responses were randomly chosen for each of the three activities described. Hence we would have 45 responses in all. Following Winer (1971), the following two source tables could be constructed, treating teacher first as a random effect (the five teachers randomly selected from a large population), then as a fixed effect (the five teachers being all that we are concerned with). In either case, replications will be considered random and nested within teacher and activity.

Table 1. Teacher as a Random Effect

Source	df	E(MS)
Activity	2	$\sigma_r^2 + 3\sigma_{t\times a}^2 + 15\sigma_a^2$
Teacher	4	$\sigma_r^2 + 9\sigma_t^2$
Activity x Teacher	8	$\sigma_r^2 + 3\sigma_{t\times a}^2$
Replications	30	σ_r^2
	44	

Table 2. Teacher as a Fixed Effect

Source	df	E(MS)
Activity	2	$\sigma_r^2 + 15\sigma_a^2$
Teacher	4	$\sigma_r^2 + 9\sigma_t^2$
Activity x Teacher	8	$\sigma_r^2 + 3\sigma_{t\times a}^2$
Replications	30	σ_r^2
	44	

As is clear from these source tables, the appropriate F ratio for the test of the Activity effect is $MS_{Act}/MS_{Act\ x\ Tch}$ with 2 and 8 degrees of freedom. To reach significance at the five percent level, this F statistic would have to be at least 4.46. If this F ratio failed to reach significance, the investigator could form the ratio MS_{Act}/MS_{Rep} and refer it to the F distribution on 2 and 30 degrees of freedom. This is a valid F test under the assumption that teacher is a fixed effect. Note that since the expectation of $MS_{Act\ x\ Tch}$ is $\sigma_r^2 + 3\sigma_{t\times a}^2$ and the expectation of MS_{Rep} is just σ_r^2 the latter F ratio will usually have a smaller denominator and hence be larger than the former. Furthermore, since MS_{Rep} is based on 30 degrees of freedom, the five percent level for the latter F is only 3.32. The higher probability of reaching significance in the second

case is plausible, since the second F tests a much weaker hypothesis. The first F ratio, treating teacher as a random effect, tested the null hypothesis that in the population of teachers from which these five were drawn, average responses were the same during the three activities. The second F ratio, treating teacher as a fixed effect, tested the null hypothesis that for these five teachers average responses were the same during the three activities.

The mood rating "elated-depressed" from the data analyzed will serve to illustrate. Due to the empty cells, only 3 df were available to estimate the activity by teacher interaction. An exact least squares solution was obtained, first entering teacher, then activity, and then the interaction into the model. Results were as follows.

Table 3. Analysis of "Elated-Depressed"

Source	df	MS	Tch Random F	Tch Random p	Tch Fixed F	Tch Fixed p
Activity	2	1.8033	1.59	>.25	3.43	.05
Teacher	4	1.3651	2.60	.06	2.60	.06
Activity x Teacher	3	1.1332	2.16	.12	2.16	.12
Replications	25					

These results indicate that an effect of activity does exist for these five teachers, but there is no evidence that the effect exists for teachers in general.

It is important to note that the behavior samples obtained via this paradigm were independent from teacher to teacher, i.e. Replications were nested within all other factors. By way of contrast, consider the following hypothetical study, which also involves two random effects and one fixed effect.

Suppose that each of five teachers employed each of three activities to teach, say, arithmetic facts, and each of three raters observed each teacher using each method. The only formal difference between this study and the hypothetical study discussed above (the idealized beeper study) is that replication was nested within teacher and activity in the idealized beeper study, and replication (rater) is crossed with teacher and activity in this study. For purposes of discussion three source tables will be presented. Table 4 is the correct source table, treating both Teacher and Rater as random effects. In table 5 Teacher is treated as a fixed effect leaving Rater as the only random effect, and in table 6 Rater is treated as fixed, leaving Teacher as the only random effect.

Table 4. Teacher Random, Rater Random

Source	df	E(MS)
Activity	2	$\sigma^2_{axtxr} + 5\sigma^2_{axr} + 3\sigma^2_{axt} + 15\sigma^2_a$
Teacher	4	$3\sigma^2_{txr} + 9\sigma^2_t$
Rater	2	$3\sigma^2_{txr} + 15\sigma^2_r$
Activity x Tch	8	$\sigma^2_{axtxr} + 3\sigma^2_{axt}$
Activity x Rater	4	$\sigma^2_{axtxr} + 5\sigma^2_{axr}$
Teacher x Rater	8	$3\sigma^2_{txr}$
Act x Tch x Rater	16	σ^2_{axtxr}
	44	

Table 5. Teacher Fixed, Rater Random

Source	df	E(MS)
Activity	2	$5\sigma^2_{axr} + 15\sigma^2_a$
Teacher	4	$3\sigma^2_{txr} + 9\sigma^2_t$
Rater	2	$15\sigma^2_r$
Activity x Tch	8	$\sigma^2_{axtxr} + 3\sigma^2_{axt}$
Activity x Rater	4	$5\sigma^2_{axr}$
Teacher x Rater	8	$3\sigma^2_{txr}$
Act x Tch x Rater	16	σ^2_{axtxr}
	44	

Table 6. Teacher Random, Rater Fixed

Source	df	E(MS)
Activity	2	$3\sigma^2_{axt} + 15\sigma^2_a$
Teacher	4	$9\sigma^2_t$
Rater	2	$3\sigma^2_{txr} + 15\sigma^2_r$
Activity x Tch	8	$3\sigma^2_{axt}$
Activity x Rater	4	$\sigma^2_{axtxr} + 5\sigma^2_{axr}$
Teacher x Rater	8	$3\sigma^2_{txr}$
Act x Tch x Rater	16	σ^2_{axtxr}
	44	

A valid F ratio to test the Activity main effect must be a ratio of two mean squares such that the expectation of the numerator exceeds the expectation of the denominator by just some multiple of σ^2_a. Inspection of the above source tables reveals that no valid F test of the Activity main effect is possible in the case that Rater and Teacher are both acknowledged as random effects. $MS_{Act}/MS_{Act \times Tch}$ tests the activity effect across teachers for the given sample of raters (Table 6) and $MS_{Act}/MS_{Act \times Rater}$ tests the generalization across raters for a given sample of teachers (Table 5), but it appears that no F statistic can be formed to test both generalizations simultaneously.

Clark (1973) has pointed out in a slightly different context that the use of many teachers but only a few raters or conversely will not be efficient in a design of this kind, since in general the test of the Activity main effect will be no more sensitive than the less sensitive of these two tests of the separate random effects. Generally the variance component for Raters will be much smaller than for Teachers, hence more teachers than raters should be used for maximum power, but if interrater reliability is low the use of more raters and fewer teachers may give more power at a given cost.

Note that if either or could be shown not to differ from zero, a valid F ratio could be constructed. Tests of each of these variance components are straightforward. $MS_{Act \times Rater}/MS_{Act \times Tch \times Rater}$ is an F ratio testing and $MS_{Act \times Tch}/MS_{Act \times Tch \times Rater}$ is an F ratio testing $\sigma^2_{a \times t}$. Winer (1971) recommends that these preliminary tests be carried out at a liberal alpha level, say p=.25. If either or both F's fail to reject, the corresponding variance component may be dropped from the expressions for the expected mean squares. After this is done, mean squares which then have the same expectation may be pooled to from the denominator of the F testing σ^2_{act}.

If both $\sigma^2_{a \times r}$ and $\sigma^2_{a \times t}$ differ significantly from zero, the only remaining alternative is the use of the quasi-F ratio, denoted F'. This is equal to $(MS_{Act} + MS_{Act \times Tch \times Rater})/(MS_{Act \times Tch} + MS_{Act \times Rater})$. Examination of Table 4 shows that the expected value of the numerator exceeds the expectation of the denominator by just $15\sigma^2_a$, but due to the pooling of the mean squares the distribution of F' is not exactly an F distribution. A satisfactory approximation to the distribution of F' is an F distribution with special degrees of freedom, computed as follows. Suppose there were p activities, q teachers, and r raters -- in tables 4, 5, and 6 p = 3, q = 5, and r = 3. Then the degrees of freedom for MS_{Act} = p-1, for $MS_{Act \times Tch \times Rater}$ = (p-1)(q-1)(r-1), for $MS_{Act \times Tch}$ = (p-1)(q-1), and for $MS_{Act \times Rater}$ = (p-1)(r-1). Denote the two numerator mean squares by u and v, and their degrees of freedom by m and n, respectively. Then for F' the numerator degrees of freedom is the nearest integer to $(u + v)^2/(u^2/m + v^2/n)$. The denominator degrees of freedom are obtained in the same way, using the mean squares and degrees of freedom for the denominator.

If rating procedures are well defined, it should be possible to avoid the use of F'. Nonetheless, rater should initially be treated as a random effect, and preliminary tests should be made to avoid faulty statistical generalizations.

This discussion has been intended to alert researchers in the area of teacher behavior to the possibilities of more than one random effect in their data, and to illustrate a few of the different designs which may arise when random factors are crossed or nested. Care in designing experiments may simplify the analysis and increase the efficiency of many studies in this area.

References

Clark, Herbert H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Wilk, M. B. & Kempthorne, O. Fixed, mixed, and random models. Journal of the American Statistical Association, 1955, 50, 1144-1167.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.